Stereo Ego-motion Improvements for Robust Rover Navigation

Clark F. Olson, Larry H. Matthies, Marcel Schoppers, and Mark W. Maimone
Jet Propulsion Laboratory, California Institute of Technology
4800 Oak Grove Drive, Pasadena CA 91109-8099

Abstract

Robust navigation for mobile robots over long distances requires an accurate method for tracking the robot position in the environment. Techniques for position estimation by determining camera ego-motion from monocular or stereo sequences have been described. However, long distance navigation requires extremely high levels of robustness and a very low rate of error growth. In this paper, we describe a methodology for long-distance navigation that meets these goals. We show that a system based on only camera ego-motion estimates will accumulate errors with superlinear growth in the distance traveled, owing to increasing orientation errors. When an absolute orientation sensor is incorporated, the error growth can be reduce to linear in the distance traveled. We have tested these techniques using both extensive simulation and hundreds of real rover images and have achieved a low, linear rate of error growth.

1 Introduction

Our goal is to perform robust and accurate rover navigation autonomously over long distances to reach terrain features with known location, but not within sight. This is motivated by the high desirability for Mars rovers to autonomously navigate to science targets observed in orbital or descent imagery. Since communication with such rovers occurs only once per day, navigation errors can result in the loss of an entire day of scientific activity.

One method for improving the position estimation errors that result from dead-reckoning is to estimate the position changes from a sequence of images (stereo or monocular) taken by the rover. This process is called ego-motion estimation. Several methods for the estimation of ego-motion have been proposed using monocular sequences [1, 2, 3, 8] and stereo sequences [4, 5, 6, 9, 10]. In order for such techniques to be effective in long-distance rover navigation, the techniques must be highly robust to problems such as poor odometry, inaccurate feature matching, and outliers.

We have developed a method is capable of accurate navigation over long distances using incremental stereo ego-motion. The use of stereo information in this method has been crucial in both outlier rejection and reducing random errors that occur due to feature localization and drift in each frame. We use a maximum-likelihood formulation of motion estimation that models error in the landmark positions more accurately than a least-squares formulation, and thus yields more accurate results. Several methods for improving the robustness of stereo ego-motion are discussed, including optimized feature selection, improved motion prediction, and multiple outlier rejection mechanisms. Reusing landmarks between frames significantly improves the overall accuracy since the errors at successive estimation steps become negatively correlated.

For long-range navigation, it is important to consider the rate of error growth as the robot travels. Even a robust system will accumulate errors that grow super-linearly with the distance traveled owing to increasing orientation errors. However, the incorporation of an absolute orientation sensor, such as a compass or sun sensor, greatly improves the long-range performance, reducing the accumulated error to a linear function of the distance traveled.

We have constructed a simulator in order to evaluate changes in the ego-motion methodology with respect to navigation performance. The simulator indicates that, with our improvements, ego-motion performance with error below 0.5% of the distance traveled is potentially feasible. Experiments on hundreds of real images have achieved errors of approximately 1% of the distance traveled.

2 Motion estimation

Our motion estimation method is based upon the maximum-likelihood ego-motion formulation of

Matthies [6, 7]. This method determines the observer motion between two (or more) pairs of stereo images captured by calibrated cameras. The basic elements of the method are as follows.

Feature selection: The first step is to select landmarks for which the 3D position can be precisely measured in successive stereo pairs. The initial landmarks are selected by finding easily trackable features in the left image of the first stereo pair.

Stereo matching (1): An estimate of the 3D position of the landmarks is obtained by performing stereo matching in the initial stereo pair. The procedure uses a correlation-based search to locate the corresponding point for each of the selected landmarks. Triangulation using the known relative position between the cameras is then used to determine the position of the landmark with respect to the camera frame. This step also provides a covariance matrix that models the error in the position estimate.

Feature tracking: Landmarks are located in subsequent stereo pairs using a correlation-based search for the selected features in the left image, that is similar to stereo matching. Prior knowledge of the approximate robot motion is used to select the search space for the feature tracking.

Stereo matching (2): A second stereo matching step is performed to estimate the 3D positions of the landmarks with respect to the new camera frame. As in the previous steps, this uses a correlation-based search and triangulation is performed to estimate the position.

Motion estimation: Motion estimation is performed using Gaussian error distributions for the landmark positions, which yields better robustness than weighted least-squares minimization [6]. The maximum-likelihood estimation problem requires an iterative solution. However, convergence is fast and requires negligible computation time compared to the previous steps.

These steps are performed for each pair of consecutive stereo frames, retaining the same set of landmarks, but replenishing those that were not found or discarded. The overall motion estimate is determined as the combination of motions from each pair of frames. Figure 1 shows the steps in the process to estimate the motion between two frames.

3 Maximum-likelihood ego-motion

Given the noisy landmark positions from stereo data, we use a maximum-likelihood formulation for motion estimation. An early version of this method was given in [6]. Further details can be found in [7].

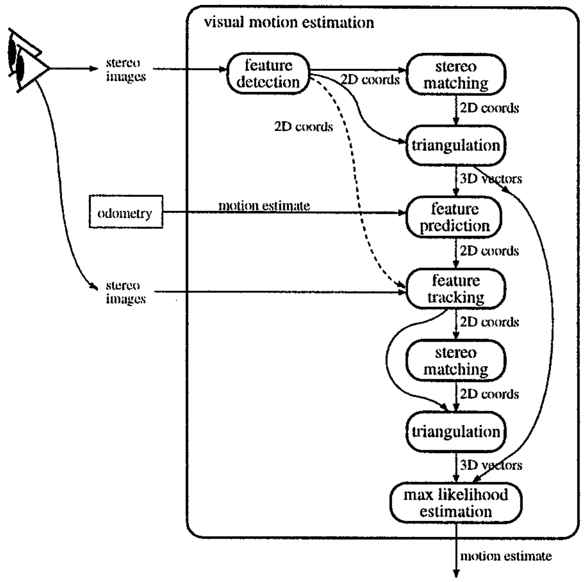


Figure 1: Steps performed for motion estimation.

Let L^b and L^a be $3 \times n$ matrices of the observed landmark positions before and after a robot motion. For each landmark we have:

$$L_i^a = RL_i^b + T + e_i, \tag{1}$$

where R and T are the rotation and translation of the robot and e combines the errors in the observed positions of the landmarks at both locations. Assume, for the moment, that the pre-move landmark positions are errorless and the post-move landmark positions are corrupted by Gaussian noise. In this case, the joint conditional probability density of the observed post-move landmark positions, given R and T, is Gaussian:

$$f(L_1^a, ..., L_n^a | R, T) \propto e^{-\frac{1}{2}\sum_{i=0}^{n} r_i^T W_i r_i}, \tag{2}$$

where $r_i = L_i^a - RL_i^b - T$ and W_i is the inverse covariance matrix of e_i. The maximum-likelihood estimate for R and T is given by minimizing the exponent $\sum_{i=0}^{n} r_i^T W_i r_i$. Note that this reduces to the least-squares solution if we let $W_i = w_i I$.

Solving for the maximum-likelihood motion estimate is a nonlinear minimization problem, which we solve through linearization and iteration. We linearize the problem by taking the first-order expansion with respect to the rotation angles. Let Θ_0 be the initial angle estimates and R_0 be the corresponding rotation matrix. The first-order expansion is:

$$L_i^a \approx R_0 L_i^b + J_i(\Theta - \Theta_0) + T + e_i, \qquad (3)$$

where J_i is the Jacobian for the ith landmark and e_i is a Gaussian noise vector with covariance $\Sigma_i = \Sigma_i^a + R_0 \Sigma_i^b R_0^T$.

We can now determine a maximum-likelihood estimate for Θ and T using $r_i = L_i^a - R_0 L_i^b - J_i(\Theta - \Theta_0) - T$ and $W_i = (\Sigma_i^a + R_0 \Sigma_i^b R_0^T)^{-1}$. Differentiating the objective function with respect to Θ and T and setting the derivatives to zero yields:

$$\left[\sum_{i=0}^{n} H_i^T W_i H_i\right] \left[\begin{array}{c} \Theta \\ T \end{array}\right] = \left[\sum_{i=0}^{n} H_i^T W_i L_i\right], \qquad (4)$$

where $H_i = [J_i \; I]$ and $L_i = L_i^a - R_0 L_i^b + J_i \Theta_0$.

After solving (4), the new motion estimate is used as an initial estimate for the next step and the process is iterated until convergence. Further details, and a technique to estimate only Θ without T, so that estimation of T can be removed from the iteration, can be found in [7].

4 Simulator

In order to test the long-range performance of the ego-motion techniques under controlled conditions, we have built a simulator that tracks randomly generated landmarks for motion estimation. The initial landmarks are generated by selecting random image location in the left image of the first (pre-move) stereo pair. The positions of the landmarks are backprojected into 3D using a random (uniformly distributed) height. Each landmark is then reprojected into the right image of the stereo pair with Gaussian noise ($\sigma = 0.3$ pixels) added in order to simulate feature matching error.

A second (post-move) stereo pair is generated using the same set of landmarks, but using camera models translated and rotated to a new position (simulating robot motion). The left image of the pair is generated by projecting the landmarks according to the new camera model and adding more Gaussian noise ($\sigma = 0.5$ pixels) in order to simulate the feature tracking error. The new image features are again backprojected into 3D (with the same heights) and reprojected into the right image of the post-move stereo pair with additional noise.

The incremental robot motion estimate is computed using the maximum-likelihood ego-motion method described above. Long-distance navigation is simulated by chaining many of the incremental moves together. At each step, the second set of landmark positions is saved for use as the initial set in the next step and new landmark positions are generated as above. When landmarks move out of the robot field of view, they are replenished with randomly positioned landmarks in the field of view.

5 Long-range error growth

Since we are interested in long-range navigation for Mars rovers, we have performed experiments examining the error growth of the stereo ego-motion techniques by applying them to a long sequence of simulated data. Our goal here is to understand the asymptotic growth of the error over long distances.

Our initial experiment considered a 500 meter traverse, with ego-motion estimates occurring ever 0.5 meters using cameras with a 45° field-of-view and 512×480 pixels. Figure 2 shows the error growth in the robot position for this experiment. It can be observed that the growth in the error is greater than linear in the distance traveled. The explanation for this is that the expected error in the orientation parameters grows approximately proportional to the square root of the distance traveled (since the overall variance is the sum of the individual variances). The overall position error grows as the sum of two terms. First, the individual position errors contribute a term that is expected to grow with the square root of the distance traveled. Second, the accumulating orientation errors contribute a term that grows as the integral of the orientation error. We thus expect a super-linear contribution from this term, which grows as $O(d^{\frac{3}{2}})$, where d is the distance traveled. The contribution from the orientation error thus dominates the overall position error.

In order to eliminate the super-linear error growth, we have examined the use of an absolute orientation sensor to provide periodic updates to the orientation estimate. For example, accelerometers can be used to provide roll and pitch information, while a compass, sun sensor, or even a panoramic camera could be used to determine the robot yaw. We have simulated such sensors as providing periodic orientation updates with Gaussian noise having zero mean and 1° standard deviation. Figure 2 shows that this results in linear error growth in the distance traveled when the orientation updates are used and, in general, the growth is much slower than when only the ego-motion estimates are used. In this experiment, the simulations indicate that

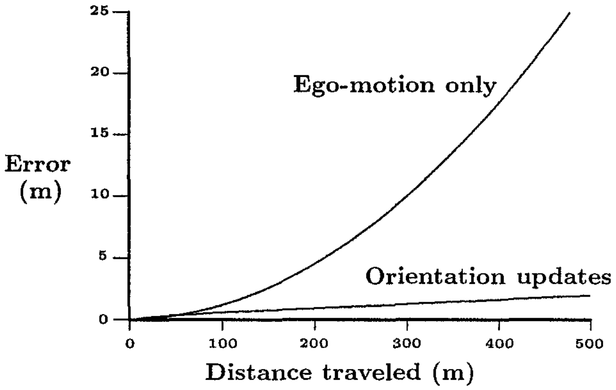


Figure 2: Expected position error as a function of distance traveled.

error less than 1% of the distance traveled is achievable with the error variances described above.

Our conclusion is that an absolute orientation sensor is critical for navigation over long distances, unless some other means is used to periodically update the robot position. If no orientation sensor is used, the robot may navigate safely over short distances. However, over long distances the increasing orientation errors will build until the position estimate is useless.

6 Robust estimation

In order to achieve accurate navigation over long distances, errors in the landmark position estimation and matching process must have a very small effect on each computed motion estimate. Tracking must be performed such that mismatches are rare. When mismatches occur, there must be mechanisms for detecting and discarding them. We describe techniques for performing these steps here, while managing the overall error buildup over time and dealing with camera roll as the robot moves.

6.1 Improved feature tracking

In many environments, it is common for the landmarks that are selected to look somewhat similar to each other and other image locations. If a large search space is necessary for each feature, incorrect matches occur frequently, since the difference in the appearance of the landmarks after the camera motion may be greater than the difference in appearance between the landmark and other image locations. For this reason, it is important to limit the search space over which we search for landmarks. Of course, we cannot limit the search space to be so small that it does not contain the correct match.

An *a priori* estimate of each landmark position is obtained using the robot odometry estimate. However, errors in the odometry incur the need for a large search window. In order to decrease the size of this search window, we estimate the robot pitch and yaw errors by first detecting a landmark near the top of the image (and thus relatively far away) using a large template window. In this case, we use a large search window, but since the landmark is also large, we are able to avoid mismatches in the image. After correcting the robot pitch and yaw estimates such that the initial landmark match is correct, we can reduce the search windows for the later correlation steps, thereby reducing the chance of a false positive.

Within the reduced search windows, our experiments have indicated that correlation using a two-resolution pyramid with decimation by a factor of four provides the best combination of speed and tracking performance.

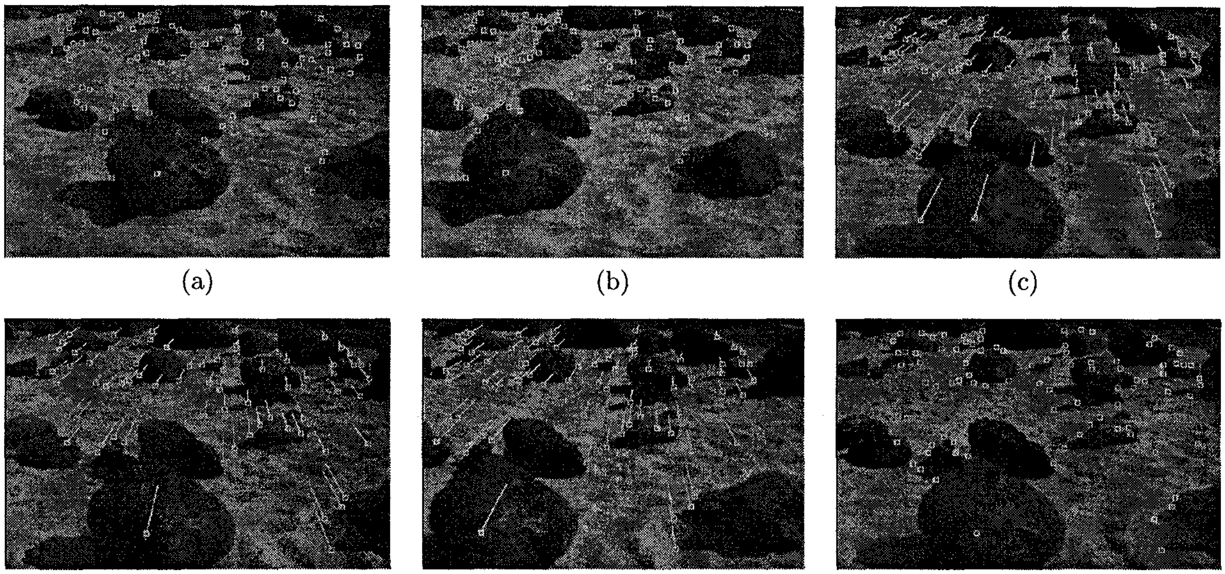
6.2 Outlier rejection

We use several methods to reject outliers in the motion estimation process. Initially, matches in both the stereo matching and feature tracking steps are eliminated if the correlation score is too low. This helps to filter out cases where a landmark is not present in the new image and cases where the change in appearance is so large that correct matching is not possible.

For each stereo match, the rays from the cameras through the image features are computed to determine if they consistent. The consistency is measured by the distance between the rays at the location of smallest separation. (If there was no error, the rays would intersect.) If this gap is not in front of the cameras, or if the projection of the gap into the image is larger than a pixel or two, the match can be rejected, since it is not geometrically feasible.

After all of the matches have been found and tracked in both stereo pairs, a rigidity test is applied to prevent gross errors. Here, we use a constraint that the landmarks must be stationary. If a landmark moves between stereo frames, the landmark is not useful for determining the robot motion. This test repeatedly rejects the landmark that appears to have moved the most, by examining the pairwise distances between the landmarks before and after the robot motion. Landmarks are rejected until all remaining deviations are small enough to be considered noise.

Finally, outlier rejection is performed within the maximum-likelihood motion estimation procedure.



(a) (b) (c) (d) (e) (f)

Figure 3: One cycle of robust feature matching. (a) Landmarks selected. (b) Landmarks matched in right image. (c) Predicted positions in next image. (d) Matched positions in left image. (e) Matched positions in right image. (f) Landmarks after replenishment.

After computing a motion estimate, the residual error for each landmark is determined. Once again, the worst matching landmarks are rejected if they have a residual greater than some threshold and the estimation is repeated.

6.3 Multi-frame tracking

Matthies [7] has shown that the errors between successive motions are negatively correlated if the same landmarks are tracked through the images. We thus expect to have lower error when the same landmarks are tracked, rather than selecting new landmarks at each step. Of course, some landmarks must be replenished at each step, since some will move out of the field-of-view and some will be rejected as outliers. However, this effect is significant, even when there is only partial overlap between the landmark sets. In our experiments, we have achieved a 27.7% reduction in navigation error when multi-frame tracking is used, rather than considering each pair of frames separately. This effect is thus useful in maintaining accurate navigation over long distances.

6.4 Camera roll

Camera roll due to traversing rough terrain is a significant problem for robots that operate outdoors. While pitch and yaw are reasonably approximated by translation of the features in the image, roll causes the features to be rotated and makes tracking significantly more difficult. Our experiments indicate that correlation scores degrade approximately linearly with the camera roll. In most terrains, camera roll of less than 10° can be tolerated without difficulty to the feature tracking.

Clearly, a robust motion estimation system for outdoor navigation must consider the effects of camera roll. The simplest solution to this problem is to ensure that image pairs are captured frequently enough that the robot does not roll by more than 10° between frames. For many systems, this solution is adequate. An alternative, for cases where large amounts of camera roll are possible, is the use of an orientation sensor, such as a gyro or accelerometer. If the approximate roll of the camera is known, then the correlation window for each landmark can be rotated to the appropriate orientation for tracking.

7 Results

These techniques have been tested on hundreds of stereo pairs, including outdoor terrain, with the robot undergoing six degree-of-freedom motion. Figure 3 shows one complete cycle of the motion estimation

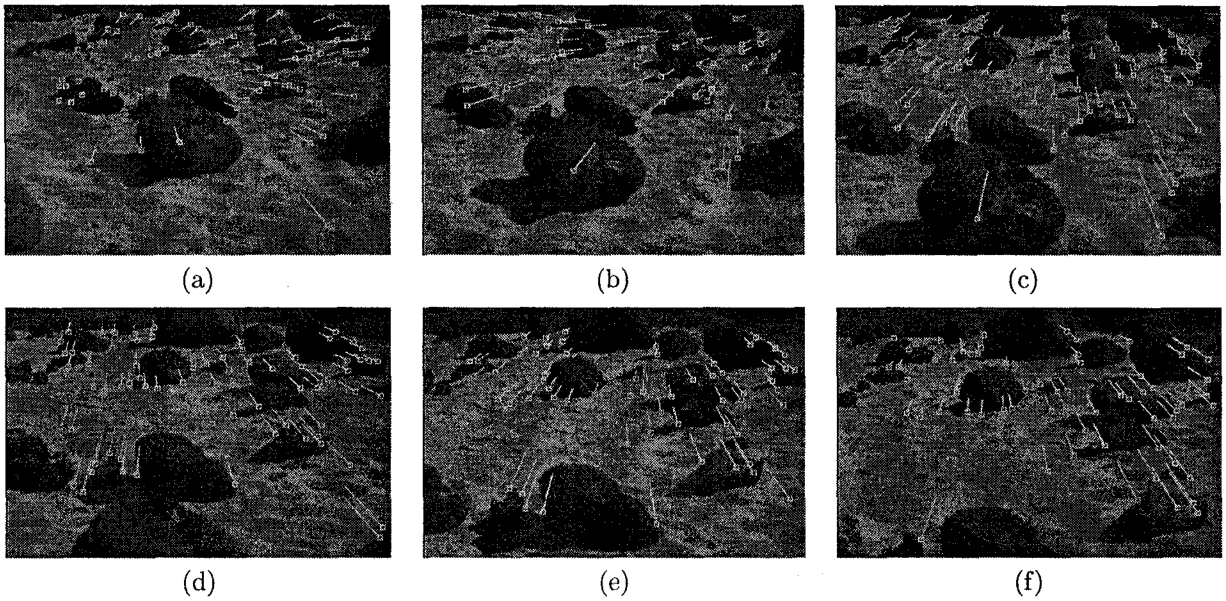


Figure 4: Several cycles of robust feature matching for ego-motion. The squares indicate the tracked landmarks and the lines show the motion of the landmark from the previous frame.

process for a simple example of forward motion. First, landmarks were selected automatically in the left image of the initial stereo pair. The matching locations were then detected in the corresponding right image. A small number of landmarks were discarded at this step due to a poor correlation score or a significant gap between the rays from the cameras. Next, the locations of the landmarks were predicted in the next image of the sequence.

After correcting for pitch and yaw error, the actual locations of the landmarks were detected in the left and right images of this image. Several landmarks were eliminated at this stage using the rigidity constraint. The remaining landmarks were used to determine the motion of the robot. Finally, the landmark set was reduced by eliminating those features that were expected to move out of the field-of-view in the next step and replenished with new landmarks.

Figure 4 shows landmark tracking for six consecutive frames of forward motion in rocky terrain. (Figure 3 corresponds to the third step in this sequence.) Despite errors in the nominal camera movements and features occurring on occluding boundaries that are difficult to track, it can be observed that the final tracking is highly robust, with no outliers in the tracking process. For this data set, the overall error was 1.3% of the distance traveled.

In order to test the performance of these techniques on extended sequences, we have applied them to imagery from a rover traverse consisting of 210 stereo pairs. This traverse was performed with a small rover and a wide field-of-view, so the cameras were close to the ground and there was considerable distortion in the appearance of close-range locations. Figure 5 shows an example of consecutive stereo pairs with 320×240 resolution. The rover traversed approximately 20 meters, taking images about every 10 centimeters. For cameras with a higher viewpoint and narrower field-of-view, the techniques could be executed less frequently. However, for this rover, small motions between stereo pairs are necessary to track the foreground landmarks. Figure 6 shows the results for this traverse. It can be observed that the ego-motion track closely follows the ground-truth from GPS, while the odometry estimate diverges from the true position. The error in this run was approximately 1.2%.

8 Summary

We have discussed techniques for improving long-range rover navigation using stereo ego-motion. An important result of our investigation is that an absolute orientation sensor is necessary to perform accurate navigation over long distances, since estimation based on ego-motion alone has error that grows super-linearly with the distance traveled. The use of an ori-

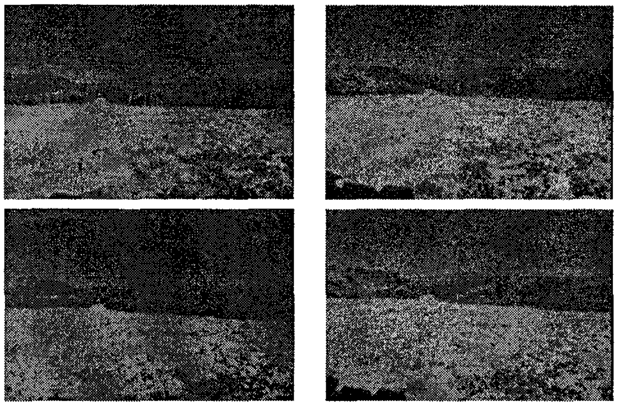


Figure 5: Stereo pairs from rover traverse sequence.

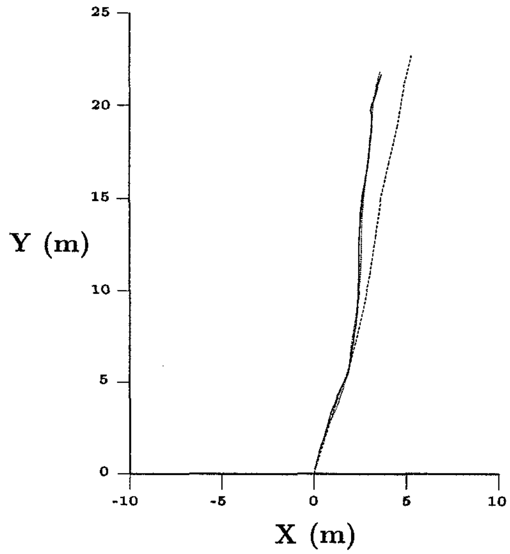


Figure 6: An extended run consisting of 210 stereo pairs. The solid line is the GPS position of the rover. The dotted line is the ego-motion estimate. The dashed line is the odometry estimate.

entation sensor reduces the error growth to linear in the distance traveled and results in a much lower error in practice. The use of stereo data was also critical to elimination of outliers and accurate motion estimation. Techniques for performing robust feature selection and tracking with outlier rejection have been developed in order to ensure accurate motion estimation at each step. We believe that this combination of techniques results in a method with greater robustness than previous techniques and that is capable of accurate motion estimation for long-distance navigation.

Acknowledgements

The research described in this paper was carried out by the Jet Propulsion Laboratory, California Institute of Technology, under a contract with the National Aeronautics and Space Administration.

References

[1] S. Chaudhuri, S. Sharma, and S. Chatterjee. Recursive estimation of motion parameters. *Computer Vision and Image Understanding*, 64(3):434–442, November 1996.

[2] D. J. Heeger and A. D. Jepson. Subspace methods for recognition rigid motion. I. algorithm and implementation. *International Journal of Computer Vision*, 7(2):95–117, January 1992.

[3] K. Kanatani. 3-d interpretation of optical flow by renormalization. *International Journal of Computer Vision*, 11(3):267–282, December 1993.

[4] S. Lacroix, A. Mallet, R. Chatila, and L. Gallo. Rover self localization in planetary-like environments. In *Proceedings of the 5th International Symposium on Artificial Intelligence, Robotics and Automation in Space*, 1999.

[5] R. Mandelbaum, G. Salgian, and H. Sawhney. Correlation-based estimation of ego-motion and structure from motion and stereo. In *Proceedings of the International Conference on Computer Vision*, volume 1, pages 544–550, 1999.

[6] L. Matthies and S. A. Shafer. Error modeling in stereo navigation. *IEEE Transactions on Robotics and Automation*, 3(3):239–248, June 1987.

[7] L. H. Matthies. *Dynamic Stereo Vision*. PhD thesis, Carnegie Mellon University, October 1989.

[8] C. Tomasi and J. Shi. Direction of heading from image deformations. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 422–427, 1993.

[9] J. Weng, P. Cohen, and N. Rebibo. Motion and structure estimation from stereo image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8(3):362–382, June 1992.

[10] Z. Zhang and O. D. Faugeras. Estimation of displacements from two 3-d frames obtained from stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(12):1141–1156, December 1992.